Medidas de Distancia (ó Distorsión)

- Una característica fundamental de los sistemas de reconocimiento (de palabra o de locutor) es la forma en que los vectores característicos son combinados y comparados con los patrones de referencia.
- Para poder realizar estas operaciones es necesario definir una **medida de distancia** entre vectores característicos.

Definición: Una distancia entre dos vectores $\mathbf{x}$ e $\mathbf{y}$ de un espacio vectorial X es una función a valores reales $d(\mathbf{x}, \mathbf{y})$ sobre el producto Cartesiano $X \times X$, que verifica las propiedades

$$(a)\ 0 \le d(\mathbf{x},\mathbf{y}) < \infty, \quad \forall\ \mathbf{x},\mathbf{y} \in X, \text{ y } d(\mathbf{x},\mathbf{y}) = 0 \text{ si y solo si } \mathbf{x} = \mathbf{y}$$

$$(b)\ d(\mathbf{x},\mathbf{y}) = d(\mathbf{y},\mathbf{x}), \ \forall\ \mathbf{x},\mathbf{y} \in X$$

$$(c)\ d(\mathbf{x},\mathbf{y}) \le d(\mathbf{x},\mathbf{z}) + d(\mathbf{y},\mathbf{z}), \ \forall\ \mathbf{x},\mathbf{y},\mathbf{z} \in X$$

- Algunas de las medidas de distancia más utilizadas son las distancias o **métricas inducidas por las normas en espacios** L_p.

Por ejemplo, si f_i, $f_i^{'}$ con $i = 1,2,\cdots,D$ son las componentes de dos vectores característicos f y $f^{'}$, pueden definirse las siguientes métricas inducidas por las normas L_p

$$d_1 = \sum_{i=1}^{D} \left| f_i - f_i^{'} \right|$$ **distancia** L_1

$$d_2 = \sqrt{\sum_{i=1}^{D} \left(f_i - f_i^{'} \right)^2}$$ **distancia Euclidea (o** L_2 **)**

- Una medida de distancia muy utilizada cuando se emplean como característica los coeficientes cepstral, que ha probado tener una muy buena performance en tareas de reconocimiento, es la **distancia Euclidea ponderada**, definida como

$$d_{2w} = \sqrt{\sum_{i=1}^{D} \left(w_i\left(c_i - c_i^{'}\right)\right)^2}$$

donde

$$w_i = \frac{1}{\sigma_i}$$

siendo σ_i^2 una estima de la varianza del i-ésimo coeficiente cepstral c_i . Aquí, los datos que son menos confiables (con mayor varianza) son pesados menos.

- Cuando se utiliza el **power cepstrum** como vector característico, a la distancia en L_2 se la denomina **distancia cepstral.** Teniendo en cuenta la identidad de Parseval, resulta

$$d_{\text{ceps}} = \sqrt{\sum_{m=-\infty}^{\infty} \left(c_m - c_m^{'}\right)^2} = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \ln\left(\left|X(\omega)\right|^2\right) - \ln\left(\left|X'(\omega)\right|^2\right) \right|^2 d\omega}$$

En la práctica, sólo se computa un número finito Q de coeficientes **power cepstral**, resultando

$$d_{\text{ceps}} \cong \sqrt{\sum_{m=1}^{Q} \left(c_m - c_m^{'}\right)^2}$$

• Una formulación más general, que tiene en cuenta la interacción entre coeficientes a traves de una matriz de covarianza es la denominada **distancia de Mahalanobis,** definida como

$$\boxed{d_M = \sqrt{\left(\overline{x} - \mu\right)^T \Sigma^{-1} \left(\overline{x} - \mu\right)}}$$

donde

$$\mu \approx \frac{1}{N_e} \sum_{i=1}^{N_e} x_i^e$$

Media de los vectores de entrenamiento x_i^e

$$\Sigma \approx \frac{1}{N_e} \sum_{i=1}^{N_e} \left(x_i^e - \mu\right)\left(x_i^e - \mu\right)^T$$

Matriz de Covarianza de los vectores de entrenamiento x_i^e

$$\overline{x} = \frac{1}{N_r} \sum_{i=1}^{N_r} x_i^r$$

Media de los vectores de reconocimiento x_i^r